10/720,618
VEHRU et al.
2530-120
2 of 2

PATENTTI- JA REKISTERIHALLITUS
NATIONAL BOARD OF PATENTS AND REGISTRATION

Helsinki 28.11.2003

E T U O I K E U S T O D I S T U S
P R I O R I T Y D O C U M E N T

Hakija Applicant	**Valtion teknillinen tutkimuskeskus** **Espoo**
Patenttihakemus nro Patent application no	**20010308**
Tekemispäivä Filing date	**16.02.2001**
Kansainvälinen luokka International class	**C12N**

Keksinnön nimitys
Title of invention

"Engineering fungi for the utilisation of L-arabinose"
(Sienen käsitteleminen geneettisesti siten, että se pystyy käyttämään L-arabinoosia)

Täten todistetaan, että oheiset asiakirjat ovat tarkkoja jäljennöksiä Patentti- ja rekisterihallitukselle alkuaan annetuista selityksestä, patenttivaatimuksista, tiivistelmästä ja piirustuksista.

This is to certify that the annexed documents are true copies of the description, claims, abstract and drawings originally filed with the Finnish Patent Office.

Pirjo Kalla
Tutkimussihteeri

Maksu 50 €
Fee 50 EUR

Maksu perustuu kauppa- ja teollisuusministeriön antamaan asetukseen 1027/2001 Patentti- ja rekisterihallituksen maksullisista suoritteista muutoksineen.

The fee is based on the Decree with amendments of the Ministry of Trade and Industry No. 1027/2001 concerning the chargeable services of the National Board of Patents and Registration of Finland.

Osoite: Arkadiankatu 6 A, P.O.Box 1160, FIN-00101 Helsinki, FINLAND
Puhelin: 09 6939 500
Telephone: + 358 9 6939 500
Telefax: 09 6939 5328
Telefax: + 358 9 6939 5328

Engineering fungi for the utilisation of L-arabinose

Field of the invention

The present invention relates to a genetically modified fungus and its use for the production of useful products such as ethanol, lactic acid, xylitol and the like from materials containing the pentose sugar L-arabinose.

Background of the invention

L-arabinose is a major constituent of plant material, L-arabinose fermentation is therefore also of potential biotechnological interest.

Fungi that can use L-arabinose and D-xylose are not necessarily good for industrial use. Many pentose utilising yeast species for example have a low ethanol tolerance, which makes them unsuitable for ethanol production. One approach would be to improve the industrial properties of these organisms. Another is to give a suitable organism the ability to use L-arabinose and D-xylose. There are pathways for D-xylose and L-arabinose, which are known to be active in bacteria. For D-xylose catabolism it is a xylose isomerase, which converts D-xylose to D-xylulose and a xylulokinase to make D-xylulose 5-phosphate. For L-arabinose catabolism the pathway consists of an isomerase, a kinase and an epimerase which convert L-arabinitol to L-ribulose, L-ribulose 5-phosphate and D-xylulose 5-phosphate, with D-xylulose 5-phosphate being an intermediate of the pentose phosphate pathway (Stryer, 1988). It has been tried to overexpress this bacterial pathway in the yeast *S. cerevisiae*, but it was not functional. The three enzymes of the L-arabinose pathway were expressed and shown to be active, however no growth on L-arabinose as a sole carbon source was reported (Sedlak and Ho, 2001). Also the expression of xylose isomerase in a fungal host was not successful (Sarthy et al. 1987, Chan et al. 1989, Kristo et al. 1989, Moes et al 1996, Schründer et al. 1996). The reason for this is not clear. There might be a species barrier, which prevents these bacterial isomerases to work in fungi. It can also be metabolic imbalances in the host, which are solved by an unknown mechanism in the donor.

There is also a hypothetical eukaryotic, i.e. fungal pathway, where L-arabinose is also converted to D-xylulose 5-phosphate, but by a different pathway (see figure 1). This pathway has been suggested to use 2 reductases, 2 dehydrogenases and a

kinase as shown (Chiang and Knight, 1961, Witteveen et al., 1989). While the genes of the bacterial pathway have been known for decades, very little is known about this hypothetical fungal pathway.

A fungal pathway for L-arabinose utilisation was described by Chiang and Knight (1961) for *Penicillium chrysogenum* and by Witteveen et al. (1989) for *Aspergillus niger*. It consists of an NADPH-linked reductase, which forms L-arabinitol, an NAD-linked dehydrogenase which forms L-xylulose, an NADPH- linked reductase which forms xylitol, an NAD-linked dehydrogenase which forms D-xylulose and a xylulokinase. The final product is D-xylulose 5-phosphate as in the bacterial L-arabinose pathway (see figure 1). This pathway was described only for filamentous fungi, but there are indications that it may also occur in yeast. Shi et al. (2000) described a mutant of *Pichia stipitis* which was unable to grow on L-arabinose. Over-expression of the NAD-linked xylitol dehydrogenase could restore the growth on L-arabinitol indicating that xylitol may be an intermediate in the L-arabinose pathway. Also yeast strains, which had L-arabinose as a sole carbon source, produced L-arabinitol and small amounts of xylitol (Dien et al., 1996), indicating that yeast might use this pathway. The capability of L-arabinose fermentation is not a common feature of yeast. Many yeast species mainly accumulate the L-arabinitol formed from L-arabinose (McMillan and Boynton 1994). Only recently yeast species were identified which were capable of L-arabinose fermentation (Dien et al., 1996).

The hypothetical fungal L-arabinose pathway has similarities to the fungal D-xylose pathway. In both pathways the pentose sugar goes through reduction and oxidation reactions where the reductions are NADPH-linked and the oxidations NAD-linked. D-xylose goes through one pair of reduction and oxidation reaction and L-arabinose goes through two pairs. The process is redox neutral but different redox cofactors, i.e. NADPH and NAD are used, which have to be separately regenerated in other metabolic pathways. In the D-xylose pathway an NADPH-linked reductase converts D-xylose into xylitol, which is then converted to D-xylulose by an NAD-linked dehydrogenase and to D-xylulose 5-phosphate by xylulokinase. The enzymes of the D-xylose pathway can all be used in the L-arabinose pathway. The first enzyme in both pathways is an aldose reductase (EC 1.1.1.21). The corresponding enzymes in *Saccharomyces cerevisiae* (Kuhn et al. 1995) and *Pichia stipitis* (Verduyn, 1985) have been characterised. They are unspecific and can use either L-arabinose or D-xylose with approximately the same rate to produce L-arabinitol or xylitol respectively. Genes coding for this enzyme are known e.g. for *Pichia stipitis*

(Amore et al., 1991), *Saccharomyces cerevisiae* (Kuhn et al., 1995, Richard et al. 1999), *Candida tenius* (Hacker et al., 1999), *Kluyveromyces lactis* (Billard et al., 1995) and *Pachysolen tannophilus* (Bolen et al., 1996).

The xylitol dehydrogenase (also known as D-xylulose reductase EC 1.1.1.9) and xylulokinase EC 2.7.1.17 are the same in the D-xylose and L-arabinose pathway of fungi. Genes for the D-xylulose reductase are known from *Pichia stipitis* (Kötter et al. 1990) *Saccharomyces cerevisiae* (Richard et al. 1999) and *Tricoderma reesei* (Wang et al. 1998). The gene for a fungal xylulokinase is only known for *Saccharomyces cerevisiae* (Ho and Chang, 1998)

Genes coding for L-arabinitol 4-dehydrogenase (EC1.1.1.12) or L-xylulose reductase (EC 1.1.1.10) are not known.

The invention aims to be able to express the pathway for L-arabinose utilisation in fungi. The hypothetical fungal pathway expressed in *Saccharomyces cerevisiae* would result in a strain, which can ferment nearly all sugars from forestry and agricultural waste to ethanol.

Summary of the invention

According to the invention, this problem is solved by genetically modifying fungus, which is characterised in that it has been transformed by a gene for L-arabinitol 4-dehydrogenase or a gene for L-xylulose reductase or both such genes.

According to the present invention, a fungus is transformed with all or some of the genes coding for the enzymes of the L-arabinose pathway, i.e. aldose reductase, L-arabinitol 4-dehydrogenase, L-xylulose reductase, D-xylulose reductase and xylulokinase. The resulting fungus is then able to utilise L-arabinose. We disclose genes for L-arabinitol dehydrogenase and L-xylulose reductase. We disclose that when a fungus as *S. cerevisiae* that is unable to utilise L-arabinose is transformed with genes for aldose reductase, L-arabinitol 4-dehydrogenase, L-xylulose reductase, D-xylulose reductase and xylulokinase, it becomes able to utilise L-arabinose. We also disclose that when a fungus, such as genetically engineered *S. cerevisiae*, that can use D-xylose but not L-arabinose is transformed with genes for L-arabinitol 4-dehydrogenase and L-xylulose reductase it can utilise L-arabinose.

By the term utilisation it is meant here for example that the organism can use L-arabinose as a carbon source or as an energy source or that it can convert L-arabinose into another compound that is a useful substance.

Brief description of the drawings

Figure 1. The hypothetical fungal and the bacterial pathway for L-arabinose utilisation.

Figure 2. L-arabinitol 4-dehydrogenase sequence (SEQ ID NO. 1): The sequence of the genomic DNA was combined with the cDNA sequences of the N-terminal and C-terminal region. The amino acid sequences in bold are from the peptide fragments of the purified protein. The intron sequence is underlined.

Figure 3. Sequence of the cDNA clone and protein sequence for the L-xylulose reductase (SEQ ID No. 2).

Detailed description of the invention

The central teaching of this invention is to demonstrate how a fungal microorganism can be genetically engineered to utilise L-arabinose. By utilization we mean that the organism can use L-arabinose as a carbon source or as an energy source or that it can convert L-arabinose into another compound that is a useful substance. Some fungi can naturally utilise L-arabinose, others cannot. It can be desirable to transfer the capacity of utilising L-arabinose to a organism lacking the capacity of L-arabinose utilisation but with other desired features, such as the ability to tolerate industrial conditions or to produce particular useful products, such as ethanol or lactic acid or xylitol. In order to transfer the capacity of L-arabinose utilisation by means of genetic engineering it is essential to know all the genes of a set of enzymes that can function together in a host cell to convert L-arabinose into a derivative, e.g. D-xylulose 5-phosphate, that the host can catabolise and so produce useful products.

One example is to genetically engineer *S. cerevisiae* to utilise L-arabinose. *S. cerevisiae* is a good ethanol producer but lacks the capacity for L-arabinose utilisation. Other examples are organisms with a useful feature but lacking at least part of a functional L-arabinose pathway.

An L-arabinose pathway believed to function in fungi is shown in the figure 1. Genes coding for the aldose reductase (EC 1.1.1.21), the D-xylulose reductase (EC 1.1.1.9) and xylulokinase (EC 2.7.1.17) are known. In order to construct a strain that can use L-arabinose by this hypothetical pathway, two additional genes would be required, i.e. genes for L-arabinitol 4-dehydrogenase (EC 1.1.1.12) and for L-xylulose reductase (EC 1.1.1.10).

L-arabinitol 4-dehydrogenase: An L-arabinitol 4-dehydrogenase was described for *Penicillium chrysogenum* and *Aspergillus niger* by Chiang and Knight (1960) and Witteveen et al (1989) respectively. This enzyme converts L-arabinitol and NAD to L-xylulose and NADH. It was also reported to have activity with NAD and adonitol (ribitol) and NAD and xylitol (Chiang and Knight, 1960).

L-xylulose reductase: The L-xylulose reductase (EC 1.1.1.10) converts xylitol and NADP to L-xylulose and NAD. Another enzyme, which has been reported to catalyse the same reaction, is the D-iditol 2-dehydrogenase (EC 1.1.1.15) (Shaw, 1956).

L-xylulose reductase was found in *Erwinia uredovora* (Doten et al, 1985), *Aspergillus niger* (Witteveen et al. 1994) and guinea pig (Hickman and Ashwell, 1959). A preparation from pigeon liver is commercially available (Sigma-Aldrich). A single subunit of the enzyme from *Aspergillus niger* has a molecular weight 32 kDa, the native enzyme an estimated weight of 250 kDa (Witteveen et al. 1994).

However, the amino acid sequences and the encoding genes are not known for any L-arabinitol dehydrogenase or L-xylulose reductase. We now disclose such genes. We also disclose that transforming these genes into a fungus that cannot utilise L-arabinose but can utilise xylose confers the ability to utilise L-arabinose upon the transformed fungus.

To identify the genes for L-arabinitol 4-dehydrogenase or L-xylulose reductase different approaches are possible and a person knowledgeable in the art might use different approaches. One approach is to purify the protein with the corresponding activity and use the information about this protein to clone the corresponding gene. This can include the proteolytic digestion of the purified protein, amino acid sequencing of the proteolytic digests and cloning a part of the gene by PCR with primers derived from the amino acid sequence. The rest of the DNA sequence can then be obtained in various ways. One way is from a cDNA library by PCR using primers from the library vector and the known part of the gene. Once the complete

sequence is known the gene can be amplified from the cDNA library and cloned into an expression vector and expressed in an heterologous host. This is a useful strategy if screening strategies or strategies, which are based on homology between sequences, are not suitable.

Another approach to clone a gene is to screen a DNA library. This is especially a good and fast procedure, when overexpression of a single gene causes a phenotype, which is easy to detect. Now that we have disclosed that transformation of a xylose-utilising fungus with genes encoding L-arabinitol dehydrogenase and L-xylulose reductase confers the ability to grow in L-arabinose, another strategy to find the genes for L-arabinitol 4-dehydrogenase and L-xylulose reductase is the following: One of the two enzymes is purified and the corresponding gene is cloned. Now all the genes of the pathway, except one, are known. In this situation a screening strategy is suitable to find the last gene of the pathway. A strain with all the genes of the pathway except one can be constructed, transformed with a DNA library, and screened for growth on L-arabinose. In this strategy one can first purify the L-arabinitol 4-dehydrogenase and then screen for the L-xylulose reductase or first purify the L-xylulose reductase and then screen for the L-arabinitol 4-dehydrogenase.

There are other ways and possibilities to clone these genes:

One can purify both enzymes and find the corresponding genes.

One can screen a DNA library or a combination of two DNA libraries to find both genes at once.

One can use other screens to find the individual genes.

One could screen for example for growth on L-xylulose to find the L-xylulose reductase and then for growth on L-arabinose or L-arabinitol to screen for L-arabinitol 4-dehydrogenase.

Other possible screens could make use of the cofactor requirements, e.g. in a screening condition which is lethal because of NADPH depletion one could screen for a L-xylulose reductase in the presence of xylitol.

One can screen existing databanks for genes with homology to genes from related protein families and test them if they have the desired enzyme activity.

For a person skilled in the art there are different ways to identify the gene, which codes for a protein with the desired enzyme activity. The methods described here illustrate our invention, but any other method known in the art may be used

Once all the genes of the L-arabinose pathway are identified, this pathway can be introduced to a new host organism, which is lacking this pathway. It is not always necessary to introduce all the genes. It might be that the host organism has already part of the pathway. For example a fungus that can utilise D-xylose might only require the enzymes that convert L-arabinitol to xylitol. Expression of L-arabinitol 4-dehydrogenase and L-xylulose reductase would then be sufficient to complete the L-arabinose pathway.

The L-arabinose pathway can be introduced to *S. cerevisiae* to generate a strain, which is a good ethanol producer and can utilise the pentoses L-arabinose and D-xylose. In such a strain the most abundant hexose and pentose sugars can be fermented to ethanol.

In Examples 3 and 5 the genes were cloned into a genetically engineered laboratory strain of *S. cerevisiae*. The same approach can be used with an industrial strain of *S. cerevisiae*, e.g. a brewer's, distiller's or baker's yeast. Industrial yeasts have process advantages such as high ethanol tolerance, tolerance of other industrial stresses and rapid fermentation. They are normally polyploid and their genetic engineering is more difficult compared to laboratory strains, but methods for their engineering are known in the art. Other yeasts unable or inefficient to utilise L-arabinose could be used as hosts, e.g. *Schizosaccaromyces pombe* or *Pichia spp.*, *Candida spp.*, *Pachysolen spp.*, *Schwanniomyces spp.*, *Arxula, spp.*, *Trichosporon spp.*, *Hansenula spp. or Yarrowia spp.* But our invention is not restricted to yeast nor even to fungi. It can be practised with any microorganism unable or inefficient to use L-arabinose.

In Examples 3 and 5 we used a TPI promoter from *S. cerevisiae* for the expression of L-arabinitol 4-dehydrogenase and the PGK promoter from *S. cerevisiae* for the expression of L-xylulose reductase. Both promoters are considered strong and constitutive. Other promoters, which are stronger or less strong, can be used. It is also not necessary to use a constitutive promoter. Inducible or repressible promoters can be used, and may have advantages, for example if a sequential fermentation of different sugars is desired.

In our example we used two plasmids for the two genes L-arabinitol 4-dehydrogenase and L-xylulose reductase. Each plasmid contained a different selection marker. These genes can also be expressed from a single plasmid with or without a selection marker or they can be integrated into the chromosomes. The selection markers were used to find successful transformations more easily and to stabilise the genetic construct. The yeast strain was transformed successively with the different genes and the transformation to *S. cerevisiae* was performed with the lithium acetate procedure (Gietz et al. 1992). This is only one method to accomplish the desired genetic construct. All the necessary genes can be transformed simultaneously or in succession. Other transformation procedures are known in the art, some being preferred for a particular host, and they can be used to achieve our invention.

In Examples 2 and 4 are disclosed the nucleotide sequences (SEQ ID NOs 1 and 2, respectively) of *T. reesei* genes encoding L-arabinitol dehydrogenase and L-xylulose reductase. These are suitable genes for practising our invention as is disclosed in Examples 5 and 6. It is well known that genes from different organisms encoding enzymes with the same catalytic activity have sequence similarities and these similarities can be exploited in many ways by those skilled in the art to clone other genes from other organisms with the same catalytic activity. Such genes are also suitable to practise our invention. It is also well known that many small variations in the nucleotide sequence of a gene do not significantly change the catalytic properties of the encoded protein. For example, many changes in nucleotide sequence do not change the amino acid sequence of the encoded protein, whereas many changes in amino acid sequence do not change the functional properties of a protein, in particular they do not prevent an enzyme from carrying out its catalytic function. We call such variations in the nucleotide sequence of DNA molecules "functionally equivalent variations" because they do not significantly change the function of the gene to encode a protein with a particular function, e.g. catalysing a particular reaction. DNA molecules that are functionally equivalent variations of the molecules defined by SEQ ID NOs 1 and 2 can be used to practise our invention.

Sometimes organisms contain genes that are not expressed under conditions that are useful in biotechnological applications. For example, although it was once generally believed that *S. cerevisiae* cannot utilise xylose and it was therefore expected that *S. cerevisae* did not contain genes encoding enzymes that would enable it to use xylose it has nevertheless been shown that *S. cerevisiae* does contain such genes

(Richard et al 1999). However, these genes are not usually expressed adequately. Thus, another aspect of our invention is to identify genes for L-arabinitol 4-dehydrogenase or L-xylulose reductase or both in a host organism itself and to cause these genes to be expressed in that same organism under conditions that are convenient for a biotechnological process, such as ethanolic fermentation of L-arabinose-containing biomass. We disclose a method of identifying candidates for such normally unexpressed genes, which is to search for similarity to SEQ ID NOs 1 and 2. A candidate gene can then be cloned in an expression vector and expressed in a suitable host and cell-free extracts of the host tested for appropriate catalytic activity as described in Examples 1 and 6. When the normally unexpressed gene has been confirmed to encode the desired enzyme, the gene can then be cloned back into the original organism but with a new promoter that causes the gene to be expressed under appropriate biotechnological conditions. This can also be achieved by genetically engineering the promoter of the gene in the intact organism.

In yet another aspect of the invention the genes encoding L-arabinitol dehydrogenase and L-xylulose reductase from a fungus, including fungi such as filamentous fungi that can have the ability to utilise L-arabinose, can now be easily identified by similarity to SEQ ID NOs 1 and 2. These genes can then be modified for example by changing their promoters to stronger promoters or promoters with different properties so as to enhance the organism's ability to utilise L-arabinose.

One embodiment of this aspect is to modify these genes (and possibly also the well known gene encoding D-xylulose reductase) to create a fungus with an enhanced capacity to produce the valuable sugar alcohols, L-arabinitol and xylitol, the latter being a useful sweetener. For example, a fungus containing aldose reductase but lacking L-arabinitol 4-dehydrogenase will convert L-arabinose to L-arabinitol and can now be created by the steps of (1) transforming the fungus with the gene for aldose reductase if it lacks this enzyme and (2) deleting or disrupting the gene for L-arabinitol 4-dehydrogenase by well known methods that utilise the sequence we disclose for this gene (SEQ ID NO 1). Similarly a fungus that contains all the enzymes of the fungal pathway for converting L-arabinose to xylitol but lacks D-xylulose reductase will convert L-arabinose into xylitol and can now be created using the information we disclose in SEQ ID NOs 1 and 2 together with information about genes for D-xylulose reductase that is already known.

A fungus may not naturally have the enzymes needed for lactic acid production, or it may produce lactic acid inefficiently. In these cases expression of the gene encoding lactate dehydrogenase (LDH) enzyme can be increased or improved in the

fungus, and a fungus can then produce lactic acid more effciently (e.g. WO 99/14335). Similarly, using methods known in the art, a fungus modified to use arabinose more efficiently as described in this invention can be further modified to produce lactic acid.

The transformed fungus of the invention may be used to produce ethanol from L-arabinose. A host fungus is transformed with genes for L-arabinitol 4-dehydrogenase, L-xylulose reductase or both. The host can be any fungus that has no or only a limited ability to use L-arabinose but is able to ferment D-xylose. For example it can be a *Saccharomyces cerevisiae* strain that has been transformed with genes enabling it to ferment D-xylose. The genes for L-arabinitol 4-dehydrogenase and L-xylulose reductase can be obtained from *T. reesei*, as described in Examples 2 and 4, but other genes encoding enzymes with these catalytic activities can also be used. Such genes are now easily found, for example from microorganisms able to use L-arabinose, because the sequences disclosed as SEQ ID Nos 1 and 2 can be exploited in various ways well known in the art to clone similar genes. The methods used to transform the host fungus and to select transformants can be the same as those used in Examples 3 and 5, but other methods known in the art can be used successfully to provide a transformed fungus according to our invention.

The transformed fungus is then used to ferment a carbon source such as biomass comprising agricultural or forestry products and waste products containing L-arabinose and other fermentable sugars. The prepartion of the carbon source for fermentation and the fermentation conditions can be the same as those that would be used to ferment the same carbon source using the host fungus. However, the transformed fungus according to the invention consumes more L-arabinose than does the host fungus and produces a higher yield of ethanol on total carbohydrate than does the host fungus. It is well known that fermentation conditions, including preparation of carbon source and fermentation temperature, agitation, gas supply, nitrogen supply, pH control, amount of fermenting organism added, can be optimised according to the nature of the raw material being fermented and the fermenting microorganism. Therefore the improved performance of the transformed fungus compared to the host fungus can be further improved by optimising the fermentation conditions according to well established process engineering procedures.

Use of a transformed fungus according to the invention to produce ethanol from carbon sources containing L-arabinose and other fermentable sugars has several industrial advantages. These include a higher yield of ethanol per ton of carbon

source and a higher concentration of ethanol in the fermented material, both of which contribute to lowering the costs of producing, for example, distilled ethanol for use as fuel. Further, the polution load in waste materials from the fermentation is lowered because the L-arabinose content is lowered, so creating a cleaner process.

Lignocellulosic raw materials are very abundant in nature and offer both renewable and and cheap carbohydrate sources for microbial processing. Arabinose containing raw materials are e.g. various pectins and hemicellulosics (such as xylans) which contain mixtures of hexoses and pentoses (xylose, arabinose). Useful raw materials include by-products from paper and pulp industry such as spent liqour and wood hydrolysates, and agricultural by-products such as sugar bagasse, corn cobs, corn fiber, oat, wheat, barley and rice hulls and straw and hydrolysates thereof. Also arabanane or galacturonic acid containing polymeric materials can be utilised.

Examples:

Example 1

Purification and amino acid sequencing of the L-arabinitol 4-dehydrogenase:

Tricoderma reesei (Rut C-30) was grown in a medium containing 40 g/l L-arabinose, 2 g/l proteose peptone, 15 g/l KH_2PO_4, 5 g/l $(NH_4)_2SO_4$, 0.6 g/l Mg_2SO_4 x 7 H_2O, 0.8 g/l $CaCl_2$ x $2H_2O$ and trace elements (Mandels and Weber, 1969) at 28 C, pH 4.0 and 30% dissolved oxygen in a fermenter (Chepmap CF2000). The fermentation was stopped when the L-arabinose was about 10 g/l. The cells were harvested with a plastic mesh sieve and washed with 10 mM sodium phosphate pH 7. 500 g of the biomass was frozen in liquid nitrogen in 100 g aliquots. After thawing and sonifying with a tip sonifyer, DTT was added to a final concentration of 5 mM and the suspension centrifuged (Sorvall SS34, 40 min, 20 000 rpm). The supernatant was dialysed overnight against a 10 fold volume of buffer A: 10 mM sodium phosphate pH 7, 5 mM DTT. The retentate was then centrifuged (Sorvall SS34, 40 min, 20 000 rpm). All steps were performed at 4 °C. The crude extract had a protein content of 7 g/l and an L-arabinitol dehydrogenase activity of 0.7 nkat per mg of extracted protein. 500 ml of this crude extract was loaded to a column with 200 ml DEAE and eluted with a linear gradient from buffer A to buffer A supplemented with 100 mM NaCl. The highest activity (16 nkat/mg, 5 mg/ml protein) eluted at about 80 mM NaCl.

The L-arabinitol 4-dehydrogenase activity was measured by adding the enzyme preparation to a buffer containing 100 mM Tris HCl pH 9.0, 0.5 mM $MgCl_2$, 2 mM NAD. The reaction was then started by adding L-arabinitol (or other sugars if specified) to a final concentration of 10 mM. The activity was calculated from the changes in NADH absorbance at 340 nm. All enzyme assays were done at 37 °C in a Cobas Mira automated analyser (Roche). In the reverse reaction the activity was measured by adding the enzyme preparation to a buffer containing 200mM $NaPO_4$ pH 7.0, 0.5 mM $MgCl_2$, 200 µM NADH and 2 mM L-xylulose. The activity was calculated from the changes in NADH absorbance at 340 nm.

The partially purified enzyme was tested for activity with other sugars. No activity was found with D-arabinitol. Activity was found with L-arabinitol and adonitol (ribitol). The activity with ribitol was about 80% of the activity found with L-arabinitol. No activity with either sugar was found when NADP was used as a cosubstrate.

In the reversible reaction with L-xylulose and NADH an activity of 0.8 nkat/mg was found with 2mM L-xylulose at pH 7.0 compared to 6.4 nkat/mg with 10 mM L-arabinitol and 5 nkat/mg with 10 mM adonitol (ribitol).

600 µl of the fraction with the highest activity after the DEAE column was then run on a native PAGE (12% acrylamide, BioRad). The gel was then stained in a Zymogram staining solution containing: 200 mM TrisHCl pH 9.0, 100 mM L-arabinitol, 0.25 mM nitroblue tetrazolium, 0.06 mM phenazine metosulfate, 1.5 mM NAD.

The only band which appeared in the staining was cut out and eluted by over-night incubation in 2ml 100 mM TrisCl pH 9.0, 0.1 % SDS. It was then concentrated to about 80 µl with Centricon (Amicon).

This gave an almost pure enzyme preparation with the major band in SDS PAGE at about 38 kDa. This protein was then used for amino acid sequencing of the proteolytic digests. The results of this sequencing were the following:

Internal peptide sequences of the purified L-arabinitol 4 dehydrogenase:

1: A T G A A I S V K P N I G V F T N P K
2: Y S N T W P R
3: A F E T S A D P K
4: H D L W I S E A E P

Example 2

Cloning of the L-arabinitol 4-dehydrogenase:

Cloning a gene fragment by using the internal amino acid sequences:

The internal peptide sequences were used to design degenerative primers for PCR. The template in the first approach was genomic DNA from *Tricoderma reesei*. A sense DNA sequence corresponding to the amino acid fragment A T G A A I S V K P N I G V F T N P K (primer 5384: ARCCIAAYATHGGIGTITTYACIAAYCC) and an anti-sense DNA sequence corresponding to the amino acid fragment A F E T S A D P K (primer 5285:GGRTCIGCIGAIGTYTCRAAIGC) were used. The PCR conditions were: denaturation 30 s, 96 °C, annealing 30 s, first 2 times 37°C and then 27 times 42°C,extention 2 min at 72°C, final extention 5 min 72°C. This procedure gave a PCR product of about 1 kb. The resulting fragment of about 1kb was then cloned to a TOPO vector (Invitrogen).

This construct was then used for sequencing.

The sequence of the PCR product coded also for the remaining two peptide sequences (see figure 2).

Cloning the N and C terminus from a cDNA library:

A cDNA library in a yeast expression vector (Margolles-Clark et al. 1996) was used to clone the residual parts of the gene. In this expression vector the cDNA is located between a PGK promoter and terminator. To clone the part of the gene, which corresponds to the N-terminus of the protein a PCR reaction was carried out with the cDNA library as a template and one primer in the PGR promoter region and an antisence primer from the gene fragment of the L-arabinitol 4-dehydrogenase.

Primer of the PGK promoter region: (primer 4196: TCAAGTTCTTAGATGCTT)

Antisence primer of the gene fragment: (primer 5431: CCTTTCCTCCAAACTTGCTGG)

The part of the gene, which corresponds to the C-terminus of the protein, was cloned in a similar way with primers from the gene fragment and an antisence primer from the PGK terminator.

Antisence primer of the PGK terminator region: (primer 3900: TAGCGTAAAGGATGGGG)

Primer of the gene fragment: (primer 5430: CTGCATTGGGCCCATGAT)

The PCR conditions were as described above except the annealing was 30 times at 50°C.

The N terminus gave a PCR product of about 0.8 kb; the C terminus gave a PCR product of about 0.9 kb. The PCR products were cloned to TOPO vectors and the resulting vectors used for sequencing.

With the information of the C-terminus and the N-terminus the open reading frame was then cloned by PCR from the cDNA library. The primer for the N-terminus contained an additional EcoRI restriction site (primer 5526: AGAATTCACCATGTCGCCTTCCGCAGTC). The primer for the C-terminus contained an additional with BamHI restriction site (primer 5468: ACGGATCCTCTACCTGGTAGCACCTCA). The annealing in the PCR reaction was 30 times 60.5 °C, Otherwise the conditions were as described above. This gave a fragment of 1.1 kb, which was then cloned to a TOPO vector and used for sequencing.

Comparing the sequences derived from genomic DNA and cDNA reveals an intron of 69 base pares (see Figure 2).

The open reading frame codes for a protein with 377 amino acids and a calculated molecular weight of 39806 g/mol.

Example 3

Expression of L-arabinitol 4-dehydrogenase in *S. cerevisiae*:

From the TOPO vector the 1.1 kb EcoRI, BamHI fragment was ligated into the corresponding sites of the pYX242 vector (R&D Systems). The pYX242 is a multicopy yeast expression vector with a yeast TPI promoter and LEU2 for selection. This plasmid was then transformed to the *S.cerevisiae* strain CEN.PK2 (VW1b). The recombinant yeast cells were grown on selective medium. The intracellular proteins were then extracted from the yeast cells by vortexing with glass beads. The extract was then analysed for L-arabinitol dehydrogenase activity.

We found an L-arabinitol 4-dehydrogenase activity of 0.2 to 0.3 nkat per mg of extracted protein.

Example 4

Screening for the L-xylulose reductase:

To screen for an L-xylulose reductase a *S. cerevisiae* strain was used which contained the genes xylose reductase (aldose reductase EC1.1.1.21), L-arabinitol-4-dehydrogenase (EC 1.1.1.12), D-xylulose reductase (EC 1.1.1.9) and xylulokinase (EC 2.7.1.17). The aldose reductase, D-xylulose reductase and xylulokinase were integrated. This strain was constructed so that uracil and leucine could still be used for selection. The plasmid from example 3 with the L-arabinitol 4-dehydrogenase on a multicopy plasmid, was transformed to the strain with the integrated aldose reductase, D-xylulose reductase and xylulokinase. In this strain the uracil auxotrophy was still left for selection. A cDNA library from *T. reesei* in a yeast expression vector with uracil marker (Margolles-Clark et al. 1996) was then transformed to this strain and screened for growth on L-arabinose. For screening the transformants were first grown on glucose plates with selection. About 750 000 transformants were then replica plated to selective plates with 5% L-arabinose as a sole carbon source. Colonies, which appeared after 2 to 3 weeks, were streaked again on L-arabinose. The resulting colonies were then grown on glucose and the plasmids rescued. The plasmids were transformed to *E. coli* cells. Since both plasmids, the plasmid with the L-arabinitol 4-dehydrogenase and the plasmid from the cDNA library, contained only ampicillin resistance, we used colony PCR to identify the *E. coli* with the cDNA library plasmid. For the colony PCR we used primers of the PGK promoter and terminator region. From 4 independent clones which appeared in the L-arabinose screening a PCR product of 0.9 kb was obtained. The corresponding plasmids were then sequenced. The sequence of the cDNA is in the figure 3. The open reading frame codes for a protein with 266 amino acids with a calculated molecular weight of 28,428 Da.

Example 5

Expression of the L-xylulose reductase:

The expression vector with the L-xylulose reductase obtained in example 4 was used. It was retransformed to the strain containing the genes xylose reductase (aldose reductase EC1.1.1.21), L-arabinitol-4-dehydrogenase (EC 1.1.1.12), D-xylulose reductase (EC 1.1.1.9) and xylulokinase (EC 2.7.1.17) which was also used in the example 4. As a control the empty vector cloning vector pAJ401 was transformed instead of the vector with the L-xylulose reductase. Transformants were first grown on D-glucose plates and then streaked on plates with 5% L-arabinose as a sole carbon source. The plates contained a carbon source and selective medium leaving out uracil and leucine as required for selection (Sherman et al. 1983). On the L-arabinose plates colonies appeared after 2 to 4 weeks with the strains with L-xylulose reductase, no colonies appeared in the control.

Example 6

Expression of the L-xylulose reductase under TPI promoter:

The L-xylulose reductase was cloned by PCR, using the vector from example 5 as a template. The primers were (LXR-start EcoRI: GCCGAATTCATCATGCCTCAGCCTGTCCCCACCGCC) and (LXR-stop HindIII: CGCCAAGCTTTTATCGTGTAGTGTAACCTCCGTCAATCAC). The conditions were as in Example 2 except that the annealing temperature was 63 °C. The PCR product was digested with EcoRI and HindIII. The vector pXY212 (R&D Systems) which is an yeast expression vector with TPI promoter and contains the URA3 gene for selection was digested with EcoRI and Hind III. The PCR product was then ligated to the expression vector. The resulting vector was then transformed to the yeast strain CEN.PK2. The recombinant yeast cells were grown on selective medium. The intracellular proteins were then extracted from the yeast cells by vortexing with glass beads. The extract was then analysed for L-xylulose reductase activity. The activity was measured in a medium containing 100mM TrisCl pH 9.0, 1.6 M xylitol and 2 mM $MgCl_2$. 2 mM NADP (final concentration) was added as a start reagent. The activity was calculated from the change in NADPH absorbance at 340 nm. The assay was performed at 37° C in a Cobas Mira automated analyser (Roche). The activity was between 2 and 5 nkat per mg of extracted protein.

References

Amore, R., Kötter, P., Kuster, C., Ciriacy, M. and Hollenberg, C.P. (1991) Cloning and expression in *Saccharomyces cerevisiae* of the NAD(P)H-dependent xylose reductase-encoding gene (XYL1) from the xylose assimilating yeast *Pichia stipitis*. Gene 109, 89-97

Billard, P., Menart, S., Fleer,.R. and Bolotin-Fukuhara, M. Isolation and characterisation of the gene encoding xylose reductase from *Kluyveromyces lactis*. Gene 162, 93-97

Bolen, P.L., Hayman, G.T. and Sheperd, H.S. Sequence and analysis of an aldose reductase gene from xylose fermenting yeast *Pachysolen tannophilus*. Yeast, 12, 1367-1375

Chan, E.-C., Ueng, P.P. and Chen, L.F. (1989) Metabolism of D-xylose in *Schizosaccharomyces pombe* cloned with a xylose isomerase gene. Appl. Micrbiol. Biotechnol. 31, 524-528

Chiang, C. and Knight, S.G. (1960) Biochem. Biophys. Res. Commun. 3, 554-559

Chiang, C and Knight, S.G. (1961) L-xylose metabolism by cell free extracts of *Penicillium chrysogenum*. Biochim. Biophys. Acta 35, 454-463

Dien, B.S., Kurtzman, C.P., Saha, B.C. and Bothast, R.J. (1996) Screening for L-arabinose fermenting yeasts. Appl. Biochem. Biotechnol. 47/48, 233-242

Doten, R.C., Mortlock, R.P.: *J. Bacteriol.*, **161**, 529-533 (1985)

Gietz, D., St Jean, A., Woods, R.A., and Schiesti, R.A. (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res. 20, 1425

Hacker, B., Habenicht, A., Kiess, M. and Mattes R. Xylose utilisation: cloning and characterisation of the xylose reductase from *Candida tenius*. Biol. Chem. 380, 1395-1403

Hickman, J., Ashwell, G.: *J. Biol. Chem.*, **234**, 758-761 (1959)

Ho, N.W.Y. and Chang, S.-F. (1989) Cloning of yeast xylulokinase gene by complementation of *E. coli* and yeast mutations. Enzyme Microb. Technol. 11, 417-421

Kötter, P, Amore, R., Hollenberg, C.P. and Ciriacy, M. (1990) Isolation and characterisation of the *Pichia stipitis* xylitol dehydrogenase gene, XYL2, and construction of a xylulose-utilizing *Saccharomyces cerevisiae* transformant. Curr. Genet. 18, 493-500

Kristo, P., Saarelainen, R., Fagerström, R., Aho, S. and Korhola, M. (1996) Protein purification, and cloning and characterization of the cDNA and gene for xylose isomerase of barley. Eur. J. Biochem. 237, 240-246

Kuhn, A., van Zyl, C., van Tonder, A. and Prior. B.A. (1995) Purification and partial characterisation of an aldo-keto reductase from *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 61, 1580-1585

McMillan, J.D. and Boynton (1994) Arabinose utilisation by xylose fermenting yeasts and fungi

Appl. Biochem. Biotechnol. 45/46, 596-584

Mandels M, Weber J (1969) The production of cellulases. Adv Chem Ser 95,391-414.

Moes, C.J., Pretorius, I.S. and van Zyl, W.H. (1996) Cloning and expression of the clostridiumthermosulfurogenes d-xylose isomerase gene (xylA) in *Saccharomyces cerevisiae*. Biotechnology letters 18, 269-274

Margolles-Clark E., Tenkanen, M., Nakari-Setälä, T. and Penttilä, M. (1996) Cloning of genes encoding alpha-L-arabinofuranosidase and beta-xylosidase from Trichoderma reesei by expression in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 62, 3840-3846

Richard, P., Toivari, M.H. and Penttilä, M. (1999) Evidence that the gene YLR070c of *Saccharomyces cerevisiae* encodes a xylitol dehydrogenase. FEBS Letters 457, 135-138

Sarthy, A.V., McConaughy, B.L., Lobo, Z, Sundstrom, J.A., Furlong, C.E. and Hall, B.D. (1987) Expression of the *Escheria coli* xylose isomerase gene in *Saccharomyces cerevisiae.* Appl. Environ. Microbiol. 53, 1996-2000

Schründer, J., Gunge, N. and Meinhardt, M. (1996) Extranuclear expression of the bacterial xylose isomerase (xylA) and the UDP-glucose dehydrogenase (hasB)

genes in yeast with *Kluyveromyces lactis* linear killer plasmids as vectors. Current Microbiology 33, 323-330

Sedlac, M. and Ho, N.W.Y. (2001) Expression of *E. coli* araBAD operon encoding enzymes for metabolizing L-arabinose in *Saccharomyces cerevisae*. Enzyme. Microb. Technol. 28, 16-24

Shaw, D.R.D.: *Biochem. J.*, **64**, 394-405 (1956)

Sherman, F., Fink, G. and Hicks, J.B. (1983) Methods in yeast genetics. A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shi, N.Q., Prahl, K., Hendrick, J., Cruz, J., Lu, P., Cho, J.Y., Jones, S. and Jeffries, T. (2000) Characterization and complementation of a *Pichia stipitis* mutant unable to grow on D-xylose or L-arabinose. Appl Biochem Biotechnol. 84-86:201-16

Stryer, L. (1988) Biochemistry, Freeman/ New York

Verduyn, C., van Kleef, R., Frank, J., Schreuder, J.H., van Dijken, J.P. and Scheffers, W.A. (1985) Properties of the NAD(P)H-dependent xylose reductase from the xylose-fermenting yeast *Pichia stipitis*

Biochem. J. 226, 668-677

Wang, T., Penttilä, M., Gao, P., Wang, C. and Zhong, L. (1998) Isolation and Identification of xylitol dehydrogenase gene from *Trichoderma reesei*. Chin. J. Biotechnol. 14, 179-185

Witteveen, C.F.B., Bunsink, R., van de Vondervoort, P., Dijkema, C., Swart, K. and Visser, J. (1989) L-arabinose and D-xylose catabolism in *Aspergillus niger*. J. Gen Microbiol 135, 2163-2171

Witteveen, C.F.B., Weber, F., Busink, R. and Visser, J. (1994) Isolation and characterisation of two xylitol dehydrogenases from *Aspergillus niger*. Microbiol. 140, 1679-1685

Claims

1. An isolated DNA molecule, **characterised** in that it comprises a gene coding for a L-arabinitol 4-dehydrogenase.

2. An isolated DNA molecule according to claim 1, **characterised** in that it has the sequence of SEQ ID NO. 1 or a functionally equivalent variants thereof.

3. An isolated DNA molecule, **characterised** in that it comprises a gene coding for a L-xylulose reductase.

4. An isolated DNA molecule according to claim 3, **characterised** in that it has the sequence SEQ ID NO. 2 or a functionally equivalent variants thereof.

5. A genetically modified fungus, **characterised** in that it has been transformed by a gene for L-arabinitol 4-dehydrogenase of claim 1 or 2 or a gene for L-xylulose reductase of claim 3 or 4 or both such genes.

6. A genetically modified fungus, **characterised** in that the expression or activity encoded by a gene for L-arabinitol 4-dehydrogenase of claim 1 or 2 or a gene for L-xylulose reductase of claim 3 or 4 or both such genes has been modified.

7. A genetically modified fungus, **characterised** in that it expresses one or more genes of the fungal L-arabinose pathway at a higher level than does the corresponding unmodified fungus under the same conditions and is able to utilise L-arabinose faster than the corresponding untransformed fungus.

8. A genetically modified fungus according to claim 7, **characterised** in that it has been transformed by a gene for L-arabinitol 4-dehydrogenase of claim 1 or 2 or a gene for L-xylulose reductase of claim 3 or 4 or both such genes.

9. A genetically modified fungus according to claim 7, **characterised** in that the expression or activity encoded by a gene for L-arabinitol 4-dehydrogenase of claim 1 or 2 or a gene for L-xylulose reductase of claim 3 or 4 or both such genes has been modified.

10. A genetically modified fungus according to any of claims 5 to 9, **characterised** in that it has an improved ability to utilise L-arabinose for growth.

11. A genetically modified fungus according to any of the claims 5 to 10, characterised in that it produces useful products.

12. A genetically modified fungus according to any of claims 5 to 11, **characterised** in that it produces intermediates in the pentose phosphate pathway or in the fungal L-arabinose pathway, or ethanol, lactic acid, xylitol or the like.

13. A genetically modified fungus according to any of claims 5 to 12, **characterised** in that it the fungus is a yeast.

14. A genetically modified fungus according to claim 13, **characterised** in that the yeast is a strain of *Saccharomyces* species, *Schizosaccharomyces* species, *Kluveromyces* species, *Pichia* species, *Candida* species or *Pachysolen* species.

15. A genetically modified fungus according to claim 14, **characterised** in that the strain is a genetically engineered strain of *S. cerevisiae*.

16. A genetically modified fungus according to any of claims 5 to 12, **characterised** in that the fungus is a filamentous fungus.

17. A genetically modified fungus according to claim 16, **characterised** in that the strain is a genetically engineered strain of *Aspergillus* species, *Trichoderma* species, *Neurospora* species, *Fusarium* species, *Penicillium* species, *Humicola* species, *Tolypocladium geodes, Trichoderma reesei (Hypocrea jecorina), Mucor species, Trichoderma longibrachiatum, Aspergillus nidulans, Aspergillus niger or Aspergillus awamori.*

18. A genetically modified fungus according to any of claims 5 to 17, **characterised** in that the fungus prior modification is able to utilise D-xylose.

19. A method of producing useful products from biomass containing L-arabinose, **characterised** in that the product is produced from said biomass by a genetically modified fungus of one of claims 5 to 18.

20. A method according to claim 19, **characterised** in that the useful product is ethanol, lactic acid or xylitol.

Abstract

A fungal microorganism can be engineered by means of genetic engineering to utilise L-arabinose. The genes of the L-arabinose pathway, which were unknown, i.e. L-arabinitol 4-dehydrogenase and L-xylulose reductase, were identified. These genes, together with the known genes of the L-arabinose pathway, form a functional pathway. This pathway can be introduced to a fungus, which is completely or partially lacking this pathway.

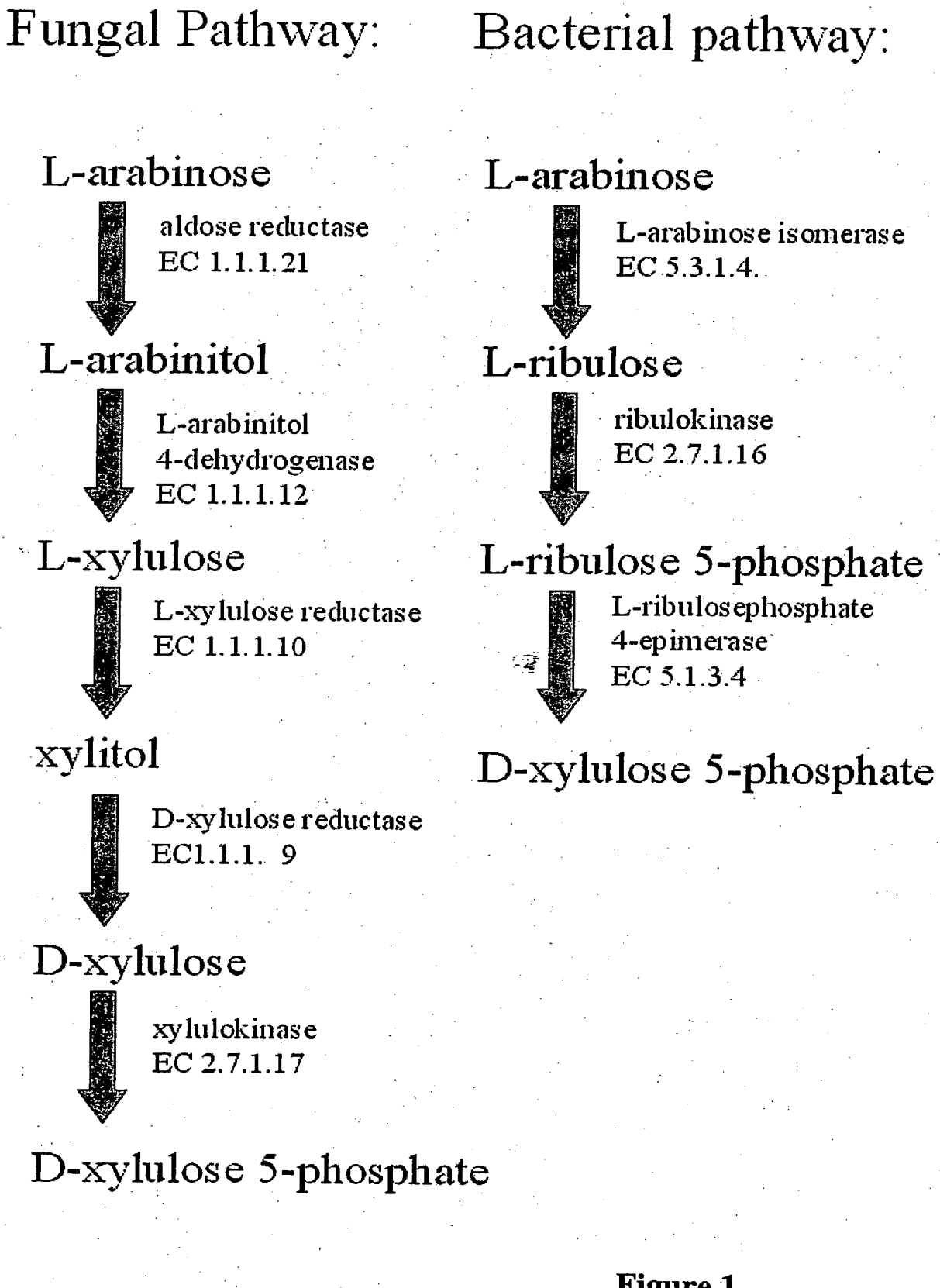
Fungal Pathway:
L-arabinose
aldose reductase
EC 1.1.1.21
L-arabinitol
L-arabinitol
4-dehydrogenase
EC 1.1.1.12
L-xylulose
L-xylulose reductase
EC 1.1.1.10
xylitol
D-xylulose reductase
EC1.1.1. 9
D-xylulose
xylulokinase
EC 2.7.1.17
D-xylulose 5-phosphate
Bacterial pathway:
L-arabinose
L-arabinose isomerase
EC 5.3.1.4.
L-ribulose
ribulokinase
EC 2.7.1.16
L-ribulose 5-phosphate
L-ribulosephosphate
4-epimerase
EC 5.1.3.4
D-xylulose 5-phosphate

Figure 1

```
1      CTCAAACGCCTTGTTCGCCGGAGACCGCGCGCATTCACAGCTCGCCATGTCGCCTTCCGC
1                                                    M  S  P  S  A

61     AGTCGATGACGCTCCCAAGGCCACAGGGGCAGCCATCTCAGTCAAGCCCAACATTGGCGT
6       V  D  D  A  P  K  A  T  G  A  A  I  S  V  K  P  N  I  G  V

121    CTTCACAAATCCAAAACATGACCTCTGGATTAGCGAAGCTGAACCCAGCGCCGATGCCGT
26      F  T  N  P  K  H  D  L  W  I  S  E  A  E  P  S  A  D  A  V

181    CAAATCTGGCGCTGATCTGAAGCCCGGCGAGGTGACCATTGCTGTCCGCAGCACTGGTAT
46      K  S  G  A  D  L  K  P  G  E  V  T  I  A  V  R  S  T  G  I

241    CTGTGGGTATGTATAACGCTTCTGTCCACAGAGCGCAAGCGCAGAGGAGCAGCATGCTGA
66      C  G

301    ACGAAATACGAATAGTTCAGATGTCCATTTCTGGCACGCCGGCTGCATTGGGCCCATGAT
68                     S  D  V  H  F  W  H  A  G  C  I  G  P  M  I

361    CGTCGAGGGCGACCACATCCTCGGCCACGAGTCTGCCGGCGAGGTCATCGCCGTCCACCC
83      V  E  G  D  H  I  L  G  H  E  S  A  G  E  V  I  A  V  H  P

421    GACTGTCAGTAGCCTCCAAATCGGCGATCGGGTTGCCATCGAGCCCAACATCATCTGCAA
103     T  V  S  S  L  Q  I  G  D  R  V  A  I  E  P  N  I  I  C  N

481    CGCGTGCGAGCCCTGCCTGACAGGTCGATACAACGGCTGCGAAAAGGTCGAGTTCCTATC
123     A  C  E  P  C  L  T  G  R  Y  N  G  C  E  K  V  E  F  L  S

541    CACGCCGCCAGTGCCCGGACCGCTGCGACGCTACGTCAACCACCCAGCCGTTTGGTGCCA
143     T  P  P  V  P  G  P  L  R  R  Y  V  N  H  P  A  V  W  C  H

601    CAAGATTGGCAACATGTCGTGGGAGAACGGCGCGCTGCTGGAGCCCCTGAGCGTGGCTCT
163     K  I  G  N  M  S  W  E  N  G  A  L  L  E  P  L  S  V  A  L

661    GGCCGGCATGCAGAGGGCCAAGGTTCAGCTCGGTGACCCCGTGCTGGTCTGCGGCGCTGG
183     A  G  M  Q  R  A  K  V  Q  L  G  D  P  V  L  V  C  G  A  G

721    TCCGATTGGATTGGTGTCAATGCTGTGCGCTGCTGCCGCCGGTGCTTGCCCGCTTGTCAT
203     P  I  G  L  V  S  M  L  C  A  A  A  A  G  A  C  P  L  V  I

781    CACAGACATTTCAGAGAGCCGTCTGGCGTTTGCAAAGGAGATCTGCCCCCGCGTCACCAC
223     T  D  I  S  E  S  R  L  A  F  A  K  E  I  C  P  R  V  T  T

841    GCACCGCATCGAGATTGGCAAGTCGGCTGAGGAAACGGCCAAAAGCATCGTCAGCTCTTT
243     H  R  I  E  I  G  K  S  A  E  E  T  A  K  S  I  V  S  S  F

901    TGGGGGCGTCGAGCCAGCCGTGACCCTGGAGTGCACCGGTGTGGAGAGCAGCATTGCAGC
263     G  G  V  E  P  A  V  T  L  E  C  T  G  V  E  S  S  I  A  A

961    GGCCATCTGGGCCAGCAAGTTTGGAGGAAAGGTCTTTGTGATCGGCGTCGGCAAGAATGA
283     A  I  W  A  S  K  F  G  G  K  V  F  V  I  G  V  G  K  N  E
```

Figure 2, cont.

```
1021    AATCAGCATTCCCTTTATGAGGGCCAGTGTACGCGAGGTCGATATCCAGCTGCAGTATCG
303       I  S  I  P  F  M  R  A  S  V  R  E  V  D  I  Q  L  Q  Y  R

1081    CTACAGCAACACCTGGCCTCGTGCCATCCGGCTCATCGAGAGCGGTGTCATCGATCTATC
323       Y  S  N  T  W  P  R  A  I  R  L  I  E  S  G  V  I  D  L  S

1141    CAAATTTGTGACGCATCGCTTCCCGCTGGAGGATGCCGTCAAGGCATTTGAGACGTCAGC
343       K  F  V  T  H  R  F  P  L  E  D  A  V  K  A  F  E  T  S  A

1201    AGATCCCAAGAGCGGCGCCATTAAGGTCATGATTCAGAGCCTGGATTGAGAGTGAGGTGC
363       D  P  K  S  G  A  I  K  V  M  I  Q  S  L  D  *

1261    TACCAGGTAGAGGTAGATAATAGATAGATGATGAAGATGGAAAGACTGCGGGCGCAAGAA


1321    TCGGGCGGATAGGGAGTTGGCTGTAATGGTTTGCAAAGCAT
```

Figure 2

```
1          CCCCATCCTTTGCATCGCCCATCATGCCTCAGCCTGTCCCCACCGCCAACAGACTCCTTG
1                                  M  P  Q  P  V  P  T  A  N  R  L  L

61         ATCTCTTCAGCTTGAAGGGCAAGGTCGTCGTCGTCACCGGCGCTTCCGGCCCTCGAGGCA
13         D  L  F  S  L  K  G  K  V  V  V  V  T  G  A  S  G  P  R  G

121        TGGGAATCGAAGCTGCCCGTGGCTGCGCCGAGATGGGCGCTGACCTCGCCATCACCTACT
33         M  G  I  E  A  A  R  G  C  A  E  M  G  A  D  L  A  I  T  Y

181        CGTCTCGCAAGGAGGGCGCGGAGAAGAACGCCGAGGAATTGACCAAGGAATACGGCGTCA
53         S  S  R  K  E  G  A  E  K  N  A  E  E  L  T  K  E  Y  G  V

241        AAGTCAAGGTGTACAAGGTCAACCAGAGCGACTACAACGATGTTGAGCGCTTTGTGAACC
73         K  V  K  V  Y  K  V  N  Q  S  D  Y  N  D  V  E  R  F  V  N

301        AGGTCGTGTCTGACTTTGGCAAGATCGATGCCTTTATTGCCAACGCCGGAGCCACAGCTA
93         Q  V  V  S  D  F  G  K  I  D  A  F  I  A  N  A  G  A  T  A

361        ATAGCGGAGTTGTTGACGGCAGCGCCAGCGATTGGGACCATGTCATCCAGGTCGACCTGA
113        N  S  G  V  V  D  G  S  A  S  D  W  D  H  V  I  Q  V  D  L

421        GCGGCACCGCATACTGCGCAAAGGCTGTTGGCGCGCACTTCAAGAAGCAGGGCCACGGCT
133        S  G  T  A  Y  C  A  K  A  V  G  A  H  F  K  K  Q  G  H  G

481        CCCTTGTCATCACAGCTTCAATGTCCGGCGACGTCGCAAACTATCCCCAGGAACAGACCT
153        S  L  V  I  T  A  S  M  S  G  H  V  A  N  Y  P  Q  E  Q  T

541        CATACAACGTCGCCAAGGCCGGTTGCATCCATCTGGCGCGGTCTCTGGCCAACGAGTGGC
173        S  Y  N  V  A  K  A  G  C  I  H  L  A  R  S  L  A  N  E  W

601        GTGATTTTGCCCGCGTCAACAGCATTTCGCCCGGTTATATCGATACCGGCCTGTCCGACT
193        R  D  F  A  R  V  N  S  I  S  P  G  Y  I  D  T  G  L  S  D

661        TCATCGACGAGAAGACGCAAGAGCTGTGGAGGAGCATGATCCCCATGGGACGAAACGGCG
213        F  I  D  E  K  T  Q  E  L  W  R  S  M  I  P  M  G  R  N  G

721        ATGCCAAGGAGCTCAAGGGCGCGTATGTATATCTGGTCAGCGACGCTAGCTCGTACACGA
233        D  A  K  E  L  K  G  A  Y  V  Y  L  V  S  D  A  S  S  Y  T

781        CGGGAGCCGATATTGTGATTGACGGAGGTTACACTACACGATAAAGAAATAATGTATTGT
253        T  G  A  D  I  V  I  D  G  G  Y  T  T  R  *

841        TAGACTATAATCAATGTGACGAACAAGATTTGTGATTAAGAAAAAAAAAAAAAAAAAAA

901        AAAACTCGAGTAATTCCGATAGA
```

Figure 3